ESC-TR-2000-071

Technical Report
1074

An Information-Theoretic Justification for Covariance Intersection and Its Generalization

M.B. Hurley

1 August 2001

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS

Prepared for the Department of the Army under Air Force
Contract F19628-00-C-0002.

This report is based on studies performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. The work was sponsored by the U.S. Army Space & Missile Defense Command under Air Force Contract F19628-00-C-0002.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Gary Tutungian
Administrative Contracting Officer
Plans and Programs Directorate
Contracted Support Management

Massachusetts Institute of Technology
Lincoln Laboratory

An Information-Theoretic Justification for Covariance Intersection and Its Generalization

M.B. Hurley
Group 33

Technical Report 1074

1 August 2001

Lexington Massachusetts

ABSTRACT

A technique for fusing Kalman filter information has been developed by Jeffrey Uhlmann, Simon Julier, et. al. that addresses the problems that arise from fusing correlated measurements. The researchers have named this technique "covariance intersection" and have presented papers on it at several robotics and control theory conferences. The technique is applicable to these areas because robotic systems often have data flowing between multiple interconnected algorithms with no guarantee that the data flowing into any algorithm are independent.

It can be shown that the covariance intersection technique is a log-linear combination of two Gaussian functions and is thus related to Chernoff information. Given this relationship, covariance intersection can be generalized to the fusion of any two probability density functions. One of the selection criteria suggested by the developers for the optimal combination of two Gaussian functions is the minimization of the determinant of the fused covariance, which is equivalent to the minimization of the Shannon information of the fused state. This equivalence justifies the selection of the determinant criterion for may applications of covariance intersection. Given the recognition of a more general rule for the covariance intersection technique, other probabilistic measures, such as the Chernoff information, may be appropriate for other fusion applications.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

1. INTRODUCTION

In the 1990s, Jeffrey Uhlmann, Simon Julier and their associates began promoting a data fusion technique that they have termed "covariance intersection" [1,2,3,4]. Their principal application of covariance intersection has been as an adjunct to Kalman filters when input data into Kalman filters are potentially highly correlated, as is often the case in complex control systems. Primary uses have focused on robotics applications, although some tests have been conducted for fusing target tracks [5].

The data fusion technique is claimed to be applicable to the fusion of sensor measurements, data estimates, or similar quantities that can be described in terms of a Gaussian probability density function. It will be shown that the covariance intersection technique is related to a more general data fusion technique that can fuse any pair of related probability density functions. This association suggests that a strict interpretation of the covariance intersection technique is that it only fuses probability density functions and not measurements, and only fuses state estimates in the sense that these estimates are represented by probability density functions. This paper presents an overview of the covariance intersection technique, the generalization of the technique, the information theoretic associations, and an example of the application of the generalized fusion technique to a simple probabilistic system.

2. COVARIANCE INTERSECTION

The covariance intersection technique is based upon the assumption that measurements or states can be described with Gaussian probability density functions. The general problem with the fusion of two probability density functions is that the two functions may have been estimated from shared measurements and therefore are coupled. The fusion of independent density functions is straightforward and widely recognized as

$$C^{-1} = A^{-1} + B^{-1}$$
$$c = C\left(A^{-1}a + B^{-1}b\right)$$

where a, b, and c are the statistical means and A, B, and C are the covariances. The appropriate application of this fusion rule is for the estimation of a probability density function that describes two independent sets of measurements assumed to be identically distributed according to a single probability density function.

Uhlmann, et al. suggest that when coupling between probability density functions is likely and the degree of coupling is unknown that a linear combination of the sigma contour

$$f_A(x) \doteq (x-a)^T A^{-1}(x-a)$$

be used, where

$$f_C(x) < \max\left(f_A(x), f_B(x)\right), \text{ and}$$

$$f_C(x) < \omega f_A(x) + (1-\omega) f_B(x). \tag{1}$$

Expansion of Equation 1 gives,

$$(x-c)^T C^{-1}(x-c) \leq \omega(x-a)^T A^{-1}(x-a) + (1-\omega)(x-b)^T B^{-1}(x-b)$$

and can be satisfied by

$$C^{-1} = \varpi A^{-1} + (1-\varpi)B^{-1}$$
$$c = C\left(\omega A^{-1}a + (1-\varpi)B^{-1}b\right)^{\cdot}$$

The remaining task is to select an appropriate value for the mixing parameter ϖ through the optimization of chosen criteria. Two criteria that have been suggested include the minimization of the determinant of the fused covariance, and the minimization of the trace of the fused covariance. It will be shown that minimization of the determinant results in a minimization of the entropy of the fused density function and is the criterion suggested by information theory.

3. GENERALIZED FUSION

Further justification for covariance intersection technique can be obtained by examining the equation used to calculate Chernoff information,

$$p_C(x) = \frac{p_A^{\varpi}(x) p_B^{1-\varpi}(x)}{\int_x p_A^{\varpi}(x) p_B^{1-\varpi}(x)}. \tag{2}$$

This function constructs a fused probability density function from a log-linear combination of two probability density functions, followed with a renormalization of the combination. Equation 2 becomes

$$p_C(x) = \frac{e^{-\varpi(x-a)^T A^{-1}(x-a)/2} e^{-(1-\varpi)(x-b)^T B^{-1}(x-b)/2}}{\int_{-\infty}^{\infty} e^{-\varpi(x-a)^T A^{-1}(x-a)/2} e^{-(1-\varpi)(x-b)^T B^{-1}(x-b)/2} dx}. \tag{3}$$

for Gaussian functions, where the normalization terms that appear in the numerator and denominator cancel. The exponential term,

$$-t/2 = -\left(\varpi(x-a)^T A^{-1}(x-a) + (1-\varpi)(x-b)^T B^{-1}(x-b)\right)/2$$

can be rewritten as

$$\begin{aligned} t &= x^T\left(\varpi A^{-1} + (1-\varpi)B^{-1}\right)x - \left(a^T A^{-1}\varpi + b^T B^{-1}(1-\varpi)\right)x - x^T\left(\varpi A^{-1}a + (1-\varpi)B^{-1}b\right) \\ &\quad + a^T \varpi A^{-1} a + b^T (1-\varpi) B^{-1} b. \end{aligned}$$

The terms in x can be gathered into a quadratic form with the substitution

$$\begin{aligned} C^{-1} &= \varpi A^{-1} + (1-\varpi)B^{-1} \\ c &= C\left(\omega A^{-1}a + (1-\varpi)B^{-1}b\right), \end{aligned}$$

exactly the same fusion formula as suggested by Uhlmann. The substitution of the defined variables results in

$$\begin{aligned} t &= x^T C^{-1} x - c^T C^{-1} x - x^T C^{-1} c + a^T \varpi A^{-1} a + b^T (1-\varpi) B^{-1} b \\ &= (x-c)^T C^{-1}(x-c) - c^T C^{-1} c + a^T \varpi A^{-1} a + b^T (1-\varpi) B^{-1} b. \end{aligned}$$

The terms independent of x in the above equation could be expanded, but these terms cancel in the numerator and denominator of Equation 3 and need not be considered further. The integration of the exponential terms containing x in the denominator is widely know and results in

$$p_C(x) = \frac{1}{(2\pi)^{n/2}} |C|^{-1/2} e^{-(x-c)^T C^{-1}(x-c)/2},$$

a Gaussian density function.

Thus, the covariance intersection technique selects a fused probability density function that is a log-linear combination of two initial probability density functions. The advantage obtained with Gaussian functions is that the fusion of two Gaussian functions results in a Gaussian function. This is not true in general for the fusion of density functions with the described technique. However, there are other families of density functions that possess this property, such as functions that are members of the statistical exponential families [6] of which Gaussian functions are members. Further research might show that other functions from the exponential families may be of interest for applications of the generalized covariance intersection technique.

3.1 THE MINIMIZATION CRITERION

Given the fusion rule, the next item to consider is the selection of the mixing parameter ω. One possible criterion is to minimize the Shannon information of the fused Gaussian function because Shannon information is a measure of the amount of information remaining to be extracted from a system under observation. The Shannon information of a probability density function is

$$I_S = -\int_x p_C(x) \ln(p_C(x)) dx$$

and for a Gaussian function is

$$I_S = -\frac{1}{(2\pi)^{n/2}} |C|^{-1/2} \int_{-\infty}^{\infty} e^{-(x-c)^T C^{-1}(x-c)/2} \ln\left(\frac{1}{(2\pi)^{n/2}} |C|^{-1/2} e^{-(x-c)^T C^{-1}(x-c)/2} \right) dx,$$

which expands to

$$I_S = -\frac{1}{(2\pi)^{n/2}} |C|^{-1/2} \int_{-\infty}^{\infty} e^{-(x-c)^T C^{-1}(x-c)/2} \left(\ln\left(\frac{1}{(2\pi)^{n/2}} |C|^{-1/2} \right) - (x-c)^T C^{-1} (x-c)/2 \right) dx.$$

The integration involving the logarithmic term is the integral of the Gaussian probability density function which, including the normalization constant, integrates to 1, giving

$$I_S = -\ln\left(\frac{1}{(2\pi)^{n/2}}|C|^{-1/2}\right) + \frac{1}{(2\pi)^{n/2}}|C|^{-1/2}\int_{-\infty}^{\infty}\left((x-c)^T C^{-1}(x-c)/2\right)e^{-(x-c)^T C^{-1}(x-c)/2}dx \tag{4}$$

Gradshteyn and Ryzhik [7] provide the integral solution

$$\int_0^{\infty} x^{2m} e^{-px^2} dx = \frac{(2m-1)!!}{2(2p)^m}\sqrt{\frac{\pi}{p}} \qquad [p > 0]$$

which gives the solution to the integral of the second term in Equation 3. The integration limits in Equation 4 allow for the substitution $x' = x - c$ without change to the form of the integral. Ignoring the prime symbol in the substitution, the solution to the integral of a principal axis of the covariance matrix can be written as

$$\int_{-\infty}^{\infty} qy^2 \exp\left(-qy^2/2\right)dy = \sqrt{\frac{2\pi}{q}} \quad . \tag{5}$$

Decomposition of the covariance matrix into principal components allows the integral in Equation 4 to be written as

$$\int_x \left((x-c)^T C^{-1}(x-c)/2\right)e^{-(x-c)^T C^{-1}(x-c)/2}dx = \frac{1}{2}\int_x\left(\sum_j q_j y_j^2\right)\exp(-\sum_k q_k y_k^2/2)dy \,.$$

The solution to the integral is thus

$$\frac{1}{2}\int_x\left(\sum_j q_j y_j^2\right)\exp(-\sum_k q_k y_k^2/2)dy = \frac{n}{2}(2\pi)^{n/2}|C|^{1/2}$$

and the full-space integration leads to the Shannon information being

$$I_S = \frac{n}{2} + \frac{1}{2}\ln\left((2\pi)^n|C|\right),$$

where n is the number of dimensions.

The Shannon information of a Gaussian function is therefore related to the determinant of the covariance. Minimization of the Shannon information of the fused Gaussian function is equivalent to the minimization of the determinant of the covariance C, which is accomplished through the appropriate choice of the mixing parameter ω.

A point to note is that Shannon information is a convex function for the family of Gaussian functions and therefore the maxima are at the ends of the chord for the covariance intersection technique in combination with the Shannon information criterion.[8] No local maxima are possible on the chord between a pair of Gaussian functions other than at the end points.

Weak associations between the Shannon information criterion and other suggested minimization criteria may be found. The inequality [9]

$$|C| \leq \left(\frac{1}{n}\operatorname{tr}(C)\right)^n$$

provides an indication as to why the trace operation would appear to work as a minimization criterion for covariance intersection. Hadamard's inequality also provides a second minimization criterion that would appear to work,

$$|C| \leq \prod_i C_{ii}\,.$$

The minima for these functions may not be identical to the minimum of the determinant. Justification for the use of these alternative minimization functions is not as strong as in the case of the Shannon information minimization and it could be argued that the apparent performance of these minimization criteria are due to their inequalities in relationship to the determinant minimization.

4. OTHER MINIMIZATION CRITERIA

Given the relationship between the Chernoff information equation and the covariance intersection technique, Chernoff information may be suitable for use as another minimization criterion for some applications. The applications most suitable for this purpose is the fusion of probability density functions where the estimates converge toward a fixed density function as the number of estimates increase. Chernoff's theorem as reformulated by Cover and Thomas[9] is "the best achievable exponent in the Bayesian probability of error is D^*, where

$$D^* = D\left(P_{\varpi^*} \parallel P_1\right) = D\left(P_{\omega^*} \parallel P_2\right),$$

with

$$P_{\varpi}(x) = \frac{P_1^{\varpi}(x) P_2^{1-\varpi}(x)}{\int_x P_1^{\varpi}(x) P_2^{1-\varpi}(x)}$$

and ϖ^* the value of ϖ such that

$$D\left(P_{\varpi^*} \parallel P_1\right) = D\left(P_{\omega^*} \parallel P_2\right).$$

where the relative entropy or Kullback-Leibler distance is

$$D(p \parallel q) = \int_x p(x) \log\left(\frac{p(x)}{q(x)}\right)."$$

Chernoff information can be used to select a probability density function that minimizes the Bayesian error. Minimization of the Shannon information will not in general provide the same solution as the minimum Bayesian error solution.

An overlooked assumption of covariance intersection is that the number of measurements that were used to estimate the two density functions are unknown and assumed to be equal. If the number of measurements that were used to derive the probability density functions are known, the Chernoff information can be modified to account for the difference in error probabilities between the two functions. Sanov's Theorem provides this connection:

"Let $X_1, X_2, \ldots, X_n$ be i.i.d. $\sim Q(x)$. Let $E \subseteq P$ be a set of probability distributions. Then

$$Q^n(E) = Q^n(E \cap P_n) \leq (n+1)^{|H|} 2^{-nD(P^* \| Q)},$$

where

$$P^* = \arg\min_{P \in E} D(P \| Q),$$

is the distribution in E that is closest to Q in relative entropy." [9]

By Sanov's theorem, the associated probabilities of error are

$$P_1^n(A) = 2^{-nD(P_\varpi \| P_1)},$$

and

$$P_2^m(A^c) = 2^{-mD(P_\varpi \| P_2)}.$$

and the total probability of error is

$$P_e(A^c) \cong \pi_1 2^{-nD(P_\varpi \| P_1)} + \pi_2 2^{-mD(P_\varpi \| P_2)} \approx 2^{-\min(nD(P_\varpi \| P_1), mD(P_\varpi \| P_2))}.$$

The exponential rate is determined by the worst exponent, and the maximum value of the minimum is obtained when the two terms are equal. For this application, we choose ϖ so that

$$nD(P_\varpi \| P_1) = mD(P_\varpi \| P_2).$$

Knowledge of the number of measurements that were used to estimate the probability density function modifies the selection of the fused probability density function toward the distribution with more measurements.

5. THREE-DIMENSIONAL EXAMPLE

The generalization of covariance intersection not only works for Gaussian functions, but for the simple case of a three-dimensional probability space. The two-dimensional probability space is simpler but less interesting; the Shannon information minimization criterion selects the member of the pair with the least information instead of an intermediate point between the two. Figure 1 shows contours of equal information on the probability simplex in probability space. Figure 2 shows the information contours, the log-linear chord between two points on the probability surface, and the location of the information minimum for the pair of points.

Figure 3 shows a plane in a log-probability space. The log-probability space is of interest because the intermediate probabilities between two probability vectors lie along a line. The conversion from probability space to log-probability space is

$$q_i = -\ln(p_i)$$

and the inverse operation is

$$p_i = \frac{\exp(-q_i)}{\sum_i \exp(-q_i)}$$

where normalization is required for the reverse mapping to probability space. The mapping from log space to probability space is an infinite-to-one mapping with the identity

$$p(q) = p(q + \alpha \mathbf{1}),$$

where α is any real number and the vector $\mathbf{1} = (1 \quad 1 \quad 1)$ for the three-dimensional space. Figure 3 shows the Shannon information for a plane in the three-dimensional log-probability space with the normal vector, $\mathbf{1}$. Examination of the Shannon information, shown as a projection out of the plane, reveals that the chord between two points in log-probability space can have both minima and maxima. It is possible to have up to two local maxima and up to three local minima (two at the ends) along the chord between two points for the three-dimensional space. The minima lie in three troughs located below the three gray arrows in Figure 3. The simple rule of selecting the Shannon information minimum from covariance intersection might be replaced with more complex rules in certain applications. The existence of a local maximum on the chord indicates that the two points represent probability density functions that are in conflict. In this case, one possible rule set might be to select the information maximum between the two points as the fused log probability point when the two points are local minima. For pairs with two maxima and three minima, the rule might be to select the central minimum if it is less than either of the end points, otherwise select the point with the global maximum on the line segment. Other rule sets are possible, and further research might show which sets are reasonable for different applications.

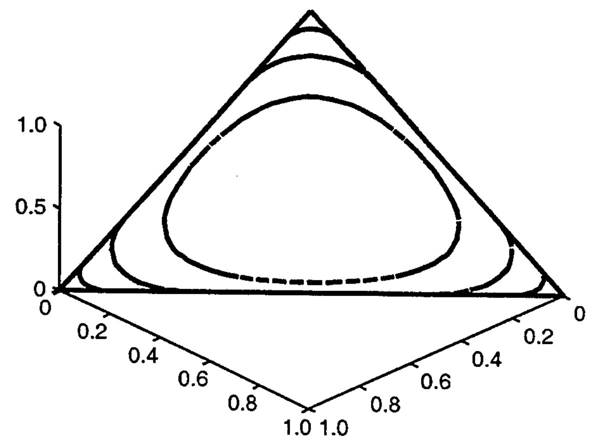

Figure 1. Shannon information contour.

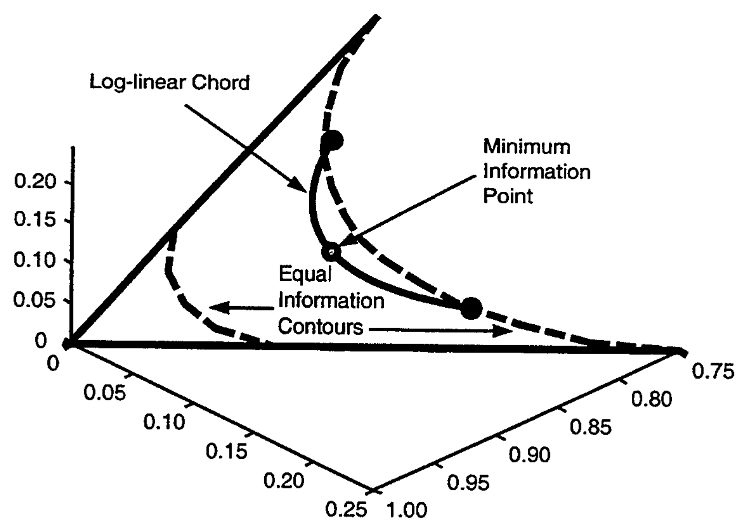

Figure 2. Log-linear chord between two points.

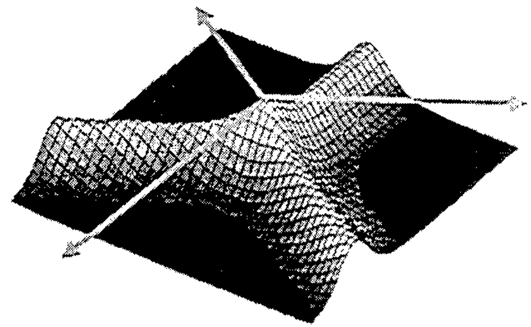

Figure 3. Information surface for a plane in log-probability space.

6. SUMMARY

It has been shown that the covariance intersection technique is related to a more general fusion technique based upon Chernoff information. The generalization shows that the covariance intersection technique finds a density function that is the log-linear combination of two initial density functions. Optimization criteria specify an optimal point on the chord between two probability density functions in a log-probability space and provide an "improved" density function. Shannon information is a natural measure for the selection criterion when knowledge of the number of measurements is unavailable and the two density functions were possibly generated from a substantial subset of common measurements.

In light of this new technique, many of the more traditional fusion techniques can be seen to be associated with estimating probability density functions appropriate for a set of measurements. With the assumption that the measurements are identical and independently distributed, additional measurements significantly restrict the set of probability density functions that describe the data. This generalized fusion of probability density functions appears to be a new technique that should be examined further to determine its range of applicability and its relationship to other known probability fusion techniques.

REFERENCES

1. J.K. Uhlmann, "General data fusion for estimates with unknown cross covariances," *Proc. SPIE*, vol. 2755, pp. 536–547 (1996).

2. J.K. Uhlmann, S.J. Julier, and M. Csorba, "Nondivergent simultaneous map-building and localization using covariance intersection," *Proc. SPIE*, vol. 3087, pp. 2–11 (1997).

3. S.J. Julier, J.K. Uhlmann, "A non-divergent estimation algorithm in the presence of unknown correlations" *Proc. Am. Control Conf.*, vol. 4, pp. 2369–2373 (1997).

4. J.K. Uhlmann, S.J. Julier, B. Kamgar-Parsi, M. Lanzagorta, and H. Shyu, "The NASA Mars Rover: a testbed for evaluating applications of covariance intersection" *SPIE Conf. Unmanned Ground Vehicle Tech.*, Orlando, FL (1999), vol. 3693, pp. 140–149.

5. J.H. Sutcliffe, D. Nicholson, and R.H. Deaves, "Data fusion for systems autonomy," *Proc. EuroFusion99*, ed. T. Windeatt and J. O'Brien (1999).

6. L.D. Brown, *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*, Hayward, CA: Institute of Mathematical Statistics (1986).

7. I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series and Products*, New York, NY: Academic Press (1980), p. 337.

8. K. Fan, "On a Theorem of Weyl concerning the eigenvalues of linear transformations II," *Proc. National Acad. Sci. U.S.*, vol. 36, pp. 31-35 (1950).

9. T.M. Cover and J.A. Thomas, *Elements of Information Theory*, New York, NY: Wiley-Interscience, (1991).

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. AGENCY USE ONLY (*Leave blank*)	2. REPORT DATE 1 August 2001	3. REPORT TYPE AND DATES COVERED Technical Report

4. TITLE AND SUBTITLE An Information-Theoretic Justification for Covariance Intersection and Its Generalization	5. FUNDING NUMBERS C—F19628-00-C-0002
6. AUTHOR(S) Dr. Michael B. Hurley	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Lincoln Laboratory, MIT 244 Wood Street Lexington, MA 02420-9108	8. PERFORMING ORGANIZATION REPORT NUMBER TR-1074
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) KMR Support Directorate, U.S. Army Space & Missile Defense Command P.O. Box 1500 Huntsville, AL 35807-3801	10. SPONSORING/MONITORING AGENCY REPORT NUMBER ESC-TR-2000-071
11. SUPPLEMENTARY NOTES None	
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.	12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

A technique for fusing Kalman filter information has been developed by Jeffrey Uhlmann, Simon Julier, et. al. that addresses the problems that arise from fusing correlated measurements. The researchers have named this technique "covariance intersection" and have presented papers on it at several robotics and control theory conferences. The technique is applicable to these areas because robotic systems often have data flowing between multiple interconnected algorithms with no guarantee that the data flowing into any algorithm are independent.

It can be shown that the covariance intersection technique is a log-linear combination of two Gaussian functions and is thus related to Chernoff information. Given this relationship, covariance intersection can be generalized to the fusion of any two probability density functions. One of the selection criteria suggested by the developers for the optimal combination of two Gaussian functions is the minimization of the determinant of the fused covariance, which is equivalent to the minimization of the Shannon information of the fused state. This equivalence justifies the selection of the determinant criterion for may applications of covariance intersection. Given the recognition of a more general rule for the covariance intersection technique, other probabilistic measures, such as the Chernoff information, may be appropriate for other fusion applications.

14. SUBJECT TERMS covariance intersection information theory Chernoff information			15. NUMBER OF PAGES 25 16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Same as Report

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by AMSI Std. 239-18
298-102